

On Translation Invariant Kernels and Screw Functions

Purushottam Kar and Harish Karnick
 Department of Computer Science and Engineering
 Indian Institute of Technology Kanpur
 {purushot,hk}@cse.iitk.ac.in

February 19, 2013

Abstract

We explore the connection between Hilbertian metrics and positive definite kernels on the real line. In particular, we look at a well-known characterization of translation invariant Hilbertian metrics on the real line by von Neumann and Schoenberg (1941). Using this result we are able to give an alternate proof of Bochner's theorem for translation invariant positive definite kernels on the real line (Rudin, 1962).

1 Introduction

We start off with a few definitions and set the notation. In the following discussion, \mathfrak{H} shall denote the real Hilbert space, small case letters x, y, \dots shall denote real numbers, boldface small letters $\mathbf{x}, \mathbf{y}, \dots$ shall denote objects in arbitrary domains. A metric D defined on some domain Ω will be said to be *Hilbertian* if there exists a map $\xi : \Omega \rightarrow \mathfrak{H}$ such that for all $\mathbf{x}, \mathbf{y} \in \Omega$, $D(\mathbf{x}, \mathbf{y}) = \|\xi(\mathbf{x}) - \xi(\mathbf{y})\|_{\mathfrak{H}}$. A symmetric real valued kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}$ will be said to be *positive definite* if for any $n \in \mathbb{N}$, any $\mathbf{x}_1, \dots, \mathbf{x}_n \in \Omega$, if we consider the matrix $G_K = [K(\mathbf{x}_i, \mathbf{x}_j)]$, then for any $c \in \mathbb{R}^n$, we have $c^\top G_K c \geq 0$. A symmetric real valued kernel $N : \Omega \times \Omega \rightarrow \mathbb{R}$ will be said to be *negative definite* if whenever $c^\top \mathbf{1} = 0$, we have $c^\top G_N c \leq 0$ (here $\mathbf{1} \in \mathbb{R}^n$ is the vector of all ones). It is easy to verify that all squared Hilbertian metrics are negative definite.

There exists a close connection between negative definite kernels and positive definite kernels as given below :

Theorem 1 ((Berg et al., 1984), Chapter 3, Lemma 2.1). *For any given kernel N over Ω , and some fixed $\mathbf{0} \in \Omega$, define a new kernel K as $K(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(N(\mathbf{x}, \mathbf{0}) + N(\mathbf{y}, \mathbf{0}) - N(\mathbf{x}, \mathbf{y}) - N(\mathbf{0}, \mathbf{0}))$. Alternatively if we have $N(\mathbf{0}, \mathbf{0}) \geq 0$ and define $K(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(N(\mathbf{x}, \mathbf{0}) + N(\mathbf{y}, \mathbf{0}) - N(\mathbf{x}, \mathbf{y}))$. Then N is negative definite iff K is positive definite.*

Using the above expression for $K(\mathbf{x}, \mathbf{y})$, it is simple to arrive at the following relation

$$-\frac{1}{2}(N(\mathbf{x}, \mathbf{x}) + N(\mathbf{y}, \mathbf{y}) - 2N(\mathbf{x}, \mathbf{y})) = K(\mathbf{x}, \mathbf{x}) + K(\mathbf{y}, \mathbf{y}) - 2K(\mathbf{x}, \mathbf{y})$$

In case N is a squared Hilbertian metric satisfying identity of the indiscernibles (i.e. $N(\mathbf{x}, \mathbf{x}) = 0$ for all $x \in \Omega$), we can obtain the following relation $N(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{x}) + K(\mathbf{y}, \mathbf{y}) - 2K(\mathbf{x}, \mathbf{y})$. This allows us to arrive at the following simple but useful corollary:

Corollary 2. *Given a distance function D over Ω satisfying identity of the indiscernibles and some fixed $\mathbf{0} \in \Omega$, define a kernel K as $K(\mathbf{x}, \mathbf{y}) = \frac{1}{2} (D^2(\mathbf{x}, \mathbf{0}) + D^2(\mathbf{y}, \mathbf{0}) - D^2(\mathbf{x}, \mathbf{y}))$. Then D is Hilbertian iff K is positive definite.*

Proof. To see that positive definiteness of K is necessary, simply invoke Theorem 1 along with the fact that since D is a metric, $D^2(\mathbf{x}, \mathbf{x}) \geq 0$ for all $x \in \Omega$ and that all squared Hilbertian metrics are negative definite. To see that is sufficient, simply invoke Mercer's theorem ((Mercer, 1909) or equivalently the discussion in (Berg et al., 1984), Chapter 3, Section 3) to confirm the existence of a mapping $\varphi : \Omega \rightarrow \mathfrak{H}$ such that $K(\mathbf{x}, \mathbf{y}) = \langle \varphi(\mathbf{x}), \varphi(\mathbf{y}) \rangle_{\mathfrak{H}}$. Thus we have $\|\varphi(\mathbf{x}) - \varphi(\mathbf{y})\|_{\mathfrak{H}}^2 = K(\mathbf{x}, \mathbf{x}) + K(\mathbf{y}, \mathbf{y}) - 2K(\mathbf{x}, \mathbf{y})$. Next, using the definition of K , we conclude that $K(\mathbf{x}, \mathbf{x}) + K(\mathbf{y}, \mathbf{y}) - 2K(\mathbf{x}, \mathbf{y}) = D^2(\mathbf{x}, \mathbf{y}) - \frac{1}{2} (D^2(\mathbf{x}, \mathbf{x}) + D^2(\mathbf{y}, \mathbf{y})) = D^2(\mathbf{x}, \mathbf{y})$ since D satisfies identity of the indiscernibles. This gives us $\|\varphi(\mathbf{x}) - \varphi(\mathbf{y})\|_{\mathfrak{H}}^2 = D^2(\mathbf{x}, \mathbf{y})$ thus proving that D is in fact a Hilbertian metric. \square

The preceding corollary and its proof give us the well known correspondence between the set of positive definite kernels and Hilbertian metrics (Schölkopf, 2000). Given any positive definite kernel K it is possible to construct a Hilbertian metric D_K and vice versa (using the polarization identity). Moreover, the preceding result holds for any arbitrary domain Ω . However if translation invariance is considered on some locally compact Abelian (LCA) group, this symmetry breaks as is demonstrated by the following discussion.

Lemma 3. *Consider a domain $\Omega \subseteq G$ where G is some LCA group. Then every translation invariant positive definite kernel K on G yields a translation invariant Hilbertian metric D .*

Proof. By translation invariance of K we have for all $\mathbf{x}, \mathbf{y}, \mathbf{t} \in G$, $K(\mathbf{x} + \mathbf{t}, \mathbf{y} + \mathbf{t}) = K(\mathbf{x}, \mathbf{y})$. Thus, for the corresponding Hilbertian metric $D^2(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{x}) + K(\mathbf{y}, \mathbf{y}) - 2K(\mathbf{x}, \mathbf{y})$, we have $D^2(\mathbf{x} + \mathbf{t}, \mathbf{y} + \mathbf{t}) = 2K(\mathbf{0}, \mathbf{0}) + K(\mathbf{x} + \mathbf{t}, \mathbf{y} + \mathbf{t}) = 2K(\mathbf{0}, \mathbf{0}) + K(\mathbf{x}, \mathbf{y}) = D^2(\mathbf{x}, \mathbf{y})$. \square

However the converse does not hold true: take for example the familiar inner product on some finite dimensional Euclidean space \mathbb{R}^d as the positive definite kernel. Its corresponding Hilbertian metric is the usual Euclidean metric $\|\cdot\|_2$. However, whereas the metric (i.e. $\|\cdot\|_2$) is translation invariant, the kernel (i.e. the inner product) is clearly not. In the sequel we shall explore this asymmetry by taking the example of the LCA group \mathbb{R} .

2 Screw Functions and Positive Definite Kernels

Von Neumann and Schoenberg (1941) initiated an investigation that resulted in a complete characterization of translation invariant Hilbertian metrics on the real line. Recall that a metric D defined on some domain Ω is said to be *Hilbertian* if there exists a map $\xi : \Omega \rightarrow \mathfrak{H}$ such that for all $\mathbf{x}, \mathbf{y} \in \Omega$, $D(\mathbf{x}, \mathbf{y}) = \|\xi(\mathbf{x}) - \xi(\mathbf{y})\|_{\mathfrak{H}}$. von Neumann and Schoenberg (1941) referred to the maps ξ that generated these metrics as *screw functions*[†]. Subsequently there were generalizations of this work to the more general class of *homogeneous Hilbertian metrics* (Fuglede, 2005). However we

[†]Von Neumann and Schoenberg (1941) also completely characterized transforms of Euclidean metrics on finite dimensional spaces that are embeddable in \mathfrak{H} , i.e. functions of the kind $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $D(\mathbf{x}, \mathbf{y}) = F(\|\mathbf{x} - \mathbf{y}\|)$ is a Hilbertian metric, thus completing a sequence of results that had previously characterized Hilbertian metric transforms for \mathfrak{H} . It is important to note that only for \mathbb{R} do these provide a complete characterization of all translation invariant Hilbertian metrics.

shall refrain from expanding our scope beyond translation invariant Hilbertian metrics on the real line. Below we restate the first main theorem of von Neumann and Schoenberg (1941):

Theorem 4 ((von Neumann and Schoenberg, 1941), Theorem 1). *All and only those translation invariant metrics on \mathbb{R} that can be expressed as the following Stieltjes integral are Hilbertian:*

$$D^2(x, y) = D^2(x - y) = \int_0^\infty \frac{\sin^2 t(x - y)}{t^2} d\gamma(t)$$

where $\gamma(t)$ is a non-decreasing function on \mathbb{R}^+ such that $\int_1^\infty t^{-2} d\gamma(t)$ exists.

Our main observation is that since translation invariant kernels all correspond to translation invariant Hilbertian metrics, the above result can be used to arrive at a characterization of translation invariant kernels on the real line as well. This result is famously known as Bochner's theorem (Rudin, 1962). As a historical note we recall that although today Bochner's theorem is taken to be the result for general locally compact Abelian groups, it was originally proved for the group of integers \mathbb{Z} by Herglotz (1911) and for the group of reals \mathbb{R} by Bochner (1933). The extension to locally compact Abelian groups is actually due to Weil (1938). Bochner was the first to recognize the key role this result plays in harmonic analysis.

However, note that it is not possible to work things the other way round, i.e. use Bochner's theorem as a starting point to arrive at the result of von Neumann and Schoenberg (1941) since not all translation invariant metrics on the real line correspond to translation invariant kernels as discussed above. This in some sense shows that, at least for the case of the real line, the result of von Neumann and Schoenberg (1941) is a more general one than Bochner's theorem and hence the two are not equivalent.

However, Bochner's theorem does not follow trivially from the above discussion and the proof progression requires overcoming a few technical hurdles. Most importantly, one is required to qualify Theorem 4 with other preconditions since the relation between translation invariant kernels and Hilbertian metrics is not that of an exact bijection.

In the subsequent discussion we will, for sake of convenience, assume that the kernel K is positive and normalized i.e. $K(x, y) \in [0, 1]$ for all x, y . This restriction will allow us to present the main ingredients of the proof without being bogged down by stray factors. We shall later relax this condition to let K assume arbitrary real values (positive as well as negative) in Appendix A. Our aim shall be to establish Bochner's theorem in the following form:

Theorem 5. *Every real valued translation invariant positive definite kernel on \mathbb{R} is the Fourier Stieltjes integral of some bounded positive Borel measure on \mathbb{R} .*

3 Proof of Theorem 5

Suppose we are given a translation invariant kernel K on the real line. We begin by constructing a distance measure D on the real line as follows (recall the relation used in the proof of Corollary 2):

$$D^2(x - y) = K(x, x) + K(y, y) - 2K(x, y),$$

where $x, y \in \mathbb{R}$. Since K is assumed to be translation invariant, D is translation invariant as well. Applying Theorem 4 on D , we arrive at the following equalities

$$\begin{aligned} 2K(0) - 2K(x, y) &= \int_0^\infty \frac{\sin^2 t(x-y)}{t^2} d\gamma(t) \\ &= \int_0^\infty \frac{1 - \cos 2t(x-y)}{2t^2} d\gamma(t) \\ &= \int_0^\infty \frac{d\gamma(t)}{2t^2} - \int_0^\infty \frac{\cos 2t(x-y)}{2t^2} d\gamma(t). \end{aligned}$$

We realize that we might have committed blasphemy by this segregation of the two parts of the integral since Theorem 4 does not assure us that $\int_0^\infty t^{-2} d\gamma(t)$ exists. It only assures us that $\int_1^\infty t^{-2} d\gamma(t)$ exists. To remedy this, we use the following simple lemma:

Lemma 6. $\int_0^\infty \frac{d\gamma(t)}{t^2} \leq 4K(0)$

Proof. von Neumann and Schoenberg (1941) characterize bounded Hilbertian metrics by proving that for any such metric D , $\lim_{t \rightarrow 0^+} \gamma(t) = \gamma(0)$ i.e. $\gamma(t)$ does not have a discrete component at 0 and that $\limsup_{t \rightarrow \infty} D^2(t) = \Theta \left(\int_0^\infty t^{-2} d\gamma(t) \right)$. More specifically, they prove

$$\frac{1}{2} \int_0^\infty t^{-2} d\gamma(t) \leq \limsup_{t \rightarrow \infty} D^2(t) \leq \int_0^\infty t^{-2} d\gamma(t).$$

In our case, we have $K(t) \in [0, 1]$ which gives us

$$D^2(t) = 2K(0) - 2K(t) \leq 2K(0).$$

This proves that our Hilbertian metric is bounded which, upon applying the result of von Neumann and Schoenberg (1941), proves the claim. \square

This justifies the segregation of the integral into two parts. Rearranging terms we get

$$\begin{aligned} K(x, y) &= K(0) - \int_0^\infty \frac{d\gamma(t)}{4t^2} + \int_0^\infty \frac{\cos 2t(x-y)}{4t^2} d\gamma(t) \\ &= C + \int_0^\infty \frac{\cos 2t(x-y)}{4t^2} d\gamma(t), \end{aligned}$$

where $C = K(0) - \int_0^\infty \frac{d\gamma(t)}{4t^2}$. Note that using Lemma 6, we have $C \geq 0$. Using a change of variables and a rescaling of the function $\gamma(\cdot)$ (which leaves all properties of $\gamma(\cdot)$ from Theorem 4 unharmed) gives us the following expression:

$$K(x, y) = C + \int_0^\infty \frac{\cos t(x-y)}{t^2} d\gamma(t)$$

for some constant $C \geq 0$.

Now we transform this Stieltjes integral into a Fourier Stieltjes integral to arrive at the canonical form of Bochner's theorem (for example as given in (Rudin, 1962)). For any $0 \leq x < \infty$, define $\alpha(x) := \int_0^x t^{-2} d\gamma(t)$ which gives us

$$K(x, y) = C + \int_0^\infty \cos t(x - y) d\alpha(t).$$

Note that $\alpha(\cdot)$ is a non-decreasing right continuous function. Now define, for any interval $S = (a, b] \subset \mathbb{R}$, the measure μ as follows

$$\mu(S) := \frac{\alpha(b) - (1 - 2 \cdot \mathbb{1}_{a < 0}) \alpha(a)}{2} + \frac{C}{2} \cdot \mathbb{1}_{0 \in (a, b]}.$$

An application of Carathéodory's extension theorem allows us to extend μ to a Borel measure over \mathbb{R} . Since $\alpha(\cdot)$ is a non decreasing function, $\int_0^\infty t^{-2} d\gamma(t) < \infty$ and $C > 0$, μ is a bounded positive measure and satisfies

$$K(x, y) = \int_{-\infty}^\infty e^{it(x-y)} d\mu(t).$$

This establishes Bochner's theorem on the real line as claimed.

4 Extensions

We note that the above characterization can also be extended to separable kernels on finite dimensional Euclidean spaces i.e. kernels which can be written as $K(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^d K_i(\mathbf{x}_i, \mathbf{y}_i)$ where $K_i(\mathbf{x}_i, \mathbf{y}_i)$ is a translation invariant kernel on the i^{th} dimension. The resulting characterization is given as the following theorem:

Theorem 7. *Every separable real valued translation invariant positive definite kernel on \mathbb{R}^d is the Fourier Stieltjes integral of some bounded positive Borel measure on \mathbb{R}^d .*

Proof. We have $K(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^d K_i(\mathbf{x}_i, \mathbf{y}_i)$ where every K_i is a translation invariant kernel. Using Theorem 5, there exist some positive bounded Borel measures μ_1, \dots, μ_n on \mathbb{R} such that for all $i = 1, \dots, d$, we have

$$K_i(\mathbf{x}_i, \mathbf{y}_i) = \int_{-\infty}^\infty e^{it(\mathbf{x}_i - \mathbf{y}_i)} d\mu_i(t).$$

Thus we have

$$\begin{aligned} K(\mathbf{x}, \mathbf{y}) &= \prod_{i=1}^d K_i(\mathbf{x}_i, \mathbf{y}_i) = \prod_{i=1}^d \int_{-\infty}^\infty e^{it_i(\mathbf{x}_i - \mathbf{y}_i)} d\mu_i(t_i) \\ &= \underbrace{\int_{-\infty}^\infty \dots \int_{-\infty}^\infty}_{d} \left(\prod_{i=1}^d e^{it_i(\mathbf{x}_i - \mathbf{y}_i)} \right) d \left(\prod_{i=1}^d \mu_i(t_i) \right) = \int_{\mathbb{R}^d} e^{i\langle \mathbf{t}, \mathbf{x} - \mathbf{y} \rangle} d\mu(\mathbf{t}), \end{aligned}$$

where $\mu = \mu_1 \times \dots \times \mu_d$ is a bounded positive product measure and we have used Fubini's theorem in the fourth step. \square

This characterization applies to some of the most widely used translation kernels (Rahimi and Recht, 2007) given below

1. Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|_2^2}{2}} = \prod_{i=1}^d e^{-\frac{(\mathbf{x}_i - \mathbf{y}_i)^2}{2}}$.
2. Laplacian kernel $K(\mathbf{x}, \mathbf{y}) = e^{-\|\mathbf{x}-\mathbf{y}\|_1} = \prod_{i=1}^d e^{-|\mathbf{x}_i - \mathbf{y}_i|}$.
3. Cauchy kernel $K(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^d \frac{2}{1 + (\mathbf{x}_i - \mathbf{y}_i)^2}$.

A Generalizing the proof for arbitrary kernels

We now assume no normalizations on K , i.e. we allow for all x, y , $K(x, y) \in \mathbb{R}$. What this lack of normalization (more specifically, the ability of the kernel to take negative values) hinders is the ability to use $K(x, y) \geq 0$ to show that $C = K(0) - \int_0^\infty \frac{d\gamma(t)}{4t^2} \geq 0$ in the proof. Thus we are left, at the end of the day, with the following characterization:

$$K(x, y) = \int_{-\infty}^{\infty} e^{it(x-y)} d\mu(t) + C,$$

where μ is a bounded positive measure and C is some (possibly negative) real constant. We first proceed by absorbing this constant into the measure μ by defining a new measure $\tilde{\mu}$. For any set $S = (a, b] \subset \mathbb{R}$, define

$$\tilde{\mu}(S) := \mu(S) + C \cdot \mathbb{1}_{0 \in (a, b]}.$$

which gives us

$$K(x, y) = \int_{-\infty}^{\infty} e^{it(x-y)} d\tilde{\mu}(t).$$

Notice that this might have caused μ to acquire a discrete component at 0 (notice that due to the property of $\gamma(\cdot)$ that $\lim_{t \rightarrow 0^+} \gamma(t) = \gamma(0)$, μ does not have a discrete component at 0 to begin with).

What we will show that unless this discrete component is non-negative, K cannot be a positive definite kernel which will establish Bochner's theorem.

This we shall show by proving that in case $C \neq 0$ (if $C = 0$ then we are done), then $\tilde{\mu}(0)$ is actually an eigenvalue of K . This will imply that for K to be positive semi-definite, $\tilde{\mu}(0) \geq 0$ which shall complete our proof.

Lemma 8. *In case the modified distribution $\tilde{\mu}$ has a discrete component $\tilde{\mu}(0)$ at 0 then $\tilde{\mu}(0)$ is an eigenvalue of K .*

Proof. For simplicity (and also without loss of generalization), assume that the domain of K is the unit interval $[-1, 1]$ so that the kernel is a function $K : [-1, 1] \times [-1, 1] \rightarrow \mathbb{R}$ (note that the kernel is still being allowed to take arbitrary real values). The purpose of proposing a compact domain is to allow us to work with constant functions $f(x) = c$ and still have $f(x) \in \ell^2$ with respect to the

Lebesgue measure. Consider the integral operator T_K on ℓ^2 corresponding to the kernel K defined as follows:

$$T_K(f)(x) = \int_{-\infty}^{\infty} f(y)K(x, y)dy.$$

We will show below that the constant function is an eigenfunction of this integral operator. Taking $f(x) = c$ we get

$$\begin{aligned} T_K(f)(x) &= \int_{-\infty}^{\infty} f(y)K(x, y)dy = \int_{-\infty}^{\infty} c \int_{-\infty}^{\infty} e^{it(x-y)} d\tilde{\mu}(t) dy \\ &= \int_{-\infty}^{\infty} ce^{itx} \left(\int_{-\infty}^{\infty} e^{-ity} dy \right) d\tilde{\mu}(t) = \int_{-\infty}^{\infty} ce^{itx} \delta(t) d\tilde{\mu}(t) \\ &= c\tilde{\mu}(0) = \tilde{\mu}(0)f(x) \end{aligned}$$

where $\delta(x)$ is the Dirac delta function and we have used Fubini's theorem in the third step. This proves the claim. \square

References

- Christian Berg, Jens Peter Reus Christensen, and Paul Russel. *Harmonic Analysis on Semigroups*. Springer-Verlag, 1984.
- Salomon Bochner. Monotone Funktionen, Stieltjessche Integrale, und harmonische Analyse. *Mathematische Annalen*, 108:378–410, 1933.
- Bent Fuglede. Spirals in Hilbert Space : With an application in information theory. *Expositiones Mathematicae*, 23:23–45, 2005.
- Gustav Herglotz. Über Potenzreihen mit positivem reellen Teil im Einheitskreis. *Leipziger Berichte*, 63:501–511, 1911.
- James Mercer. Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations. *Philosophical Transactions of the Royal Society A*, 209:415–446, 1909.
- Ali Rahimi and Benjamin Recht. Random Features for Large-Scale Kernel Machines. In *Neural Information Processing Systems*, 2007.
- Walter Rudin. *Fourier Analysis on Groups*. Interscience Publishers, New. York, 1962.
- Bernhard Schölkopf. The Kernel Trick for Distances. In *Advances in Neural Information Processing Systems*, pages 301–307, 2000.
- John von Neumann and Isaac Jacob Schoenberg. Fourier Integrals and Metric Geometry . *Transactions of the American Mathematical Society*, 50:226–251, 1941.
- André Weil. *L'intégration dans les groupes topologiques et ses applicaitons*. Gauthiers-Villars, Paris, 1938.